Article

Social Science Computer Review
1-19

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0894439319879132
journals.sagepub.com/home/ssc
SAGE

Slider Bars in Multi-Device Web Surveys

Angelica M. Maineri[1], Ivano Bison[2], and Ruud Luijkx[1,2]

Abstract

This study explores some features of slider bars in the context of a multi-device web survey. Using data collected among the students of the University of Trento in 2015 and 2016 by means of two web surveys ($N = 6{,}343$ and 4,124) including two experiments, we investigated the effect of the initial position of the handle and the presence of numeric labels on answers provided using slider bars. It emerged that the initial position of the handle affected answers and that the number of rounded scores increased with numeric feedback. Smartphone respondents appeared more sensitive to the initial position of the handle but also less affected by the presence of numeric labels resulting in a lower tendency to rounding. Yet, outcomes on anchoring were inconclusive. Overall, no relevant differences have been detected between tablet and PC respondents. Understanding to what extent interactive and engaging tools such as slider bars can be successfully employed in multi-device surveys without affecting data quality is a key challenge for those who want to exploit the potential of web-based and multi-device data collection without undermining the quality of measurement.

Keywords

slider question, sliders, multi-device surveys, web surveys, mobile surveys, survey experiment

The implementation of surveys on the web provided the opportunity to use innovative and interactive measurement tools such as slider bars—graphic elements which require respondents to drag a marker along a line to select an answer or assign a score to a stimulus. The advantage of this type of rating scale is that it offers a continuum on which respondents can freely place their answer suggesting equal intervals (Cook, Heath, Thompson, & Thompson, 2001).

While advantages and drawbacks of implementing continuous rating scales in web surveys have been discussed, visual features of slider bars, for example, the initial position of the handle, need to be investigated (DeCastellarnau, 2018). Furthermore, recent developments in mobile technologies pose new challenges to researchers, since the spread of survey completion via smartphones and tablets and the subsequent need to develop surveys suitable for small screens question the usability of graphic rating elements.

[1] Tilburg University, The Netherlands
[2] University of Trento, Italy

Corresponding Author:
Angelica M. Maineri, Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands.
Email: a.m.maineri@uvt.nl

This article investigates the usability of slider bars in multi-device web surveys. The contribution is manifold. First of all, this work expands on existing literature by exploring visual features of slider bars using two survey experiments. Secondly, the study contributes to the literature on the (unintended) mobile access to web surveys and device effects. The insights we offer may be useful for survey practitioners who want to exploit the potential of web-based data collection by using interactive tools such as slider bars, whereas ensuring that measurement quality is not endangered.

Background

Slider Bars

Slider bars are a type of continuous rating scale enabled by web surveys, in which respondents are requested to move a handle (or marker or slider) along a continuous horizontal line in order to select an answer/score (Chyung, Swanson, Roberts, & Hankinson, 2018). Slider bars differ from Visual Analogue Scales (VAS) in the movement required, which is *point and click* in VAS and *drag and drop* in slider bars (Chyung et al., 2018). Both tools offer continuous rating scales, which are considered more engaging for respondents (Funke, Reips, & Thomas, 2011) and superior in terms of measurement possibilities compared to discrete rating scales (DeCastellarnau, 2018; Roster, Lucianetti, & Albaum, 2015).

Many studies compared measurement quality and completion outcomes between continuous and discrete rating scales (see Chyung et al., 2018, for an overview). Despite concerns about completion outcomes (e.g., the amount of missing data and completion time), evidence on measurement quality gathered by web survey experiments seems to be encouraging for VAS (Couper, Tourangeau, Conrad, & Singer, 2006; Funke & Reips, 2012; Liu & Conrad, 2016). Nonetheless, findings comparing slider bars and radio buttons are less optimistic (Buskirk, Saunders, & Michaud, 2015; Maloshonok & Terentev, 2016; Roster et al., 2015). Even when no threats to data quality are detected, the measurement does not improve sensibly (Bosch, Revilla, Decastellarnau, & Weber, 2019). Moreover, slider bars are reported to perform worse than VAS in web survey experiments comparing different types of answer scales (Funke, 2016; Toepoel & Funke, 2018).

The problematic character of slider bars may be due to the peculiar visual features of these tools, and from their usability in multi-device web surveys, which this study aims to explore further.

Visual Features of Slider Bars

Among the features of slider bars that may affect responses, the most cited is the initial position of the handle, which may lead respondents to select a score on a certain portion of the bar (DeCastellarnau, 2018; Funke et al., 2011). A few studies fielding survey experiments in nonprobability online access panels seem to support this claim: the default position of the slider affected the scores (Buskirk et al., 2015; Liu & Conrad, 2019) and the amount of missing items (Buskirk et al., 2015). Toepoel and Funke (2018) found that placing the marker on the left extreme biased answers toward the left side of the bar. In an experiment on a probability-based online panel in Norway, Bosch, Revilla, Decastellarnau, and Weber (2019) found that, compared to radio buttons, slider bars performed better when the marker was at the middle rather than on the left. Another feature that may affect the performance of the slider bars is the numeric feedback and/or labeling (Roster et al., 2015), namely providing the respondent with the number corresponding to the score selected with the handle. To our knowledge, there are no studies up to date focusing on this feature on slider bars. Liu and Conrad (2016) and Couper, Tourangeau, Conrad, and Singer (2006) fielded experiments involving numeric feedback on VAS and found that numeric feedback encourages rounding.

Device Effects

The penetration of mobile Internet-enabled devices has increased tremendously over the last few years (De Bruijne & Wijnant, 2014), widening the range of devices employed to complete a survey (Callegaro, 2010). The analysis of the consequences on data quality of the (unintended) mobile access to online surveys surely constitutes one of the major challenges for survey methodologists nowadays. For an extensive review on mobile web surveys, see Couper, Antoun, and Mavletova (2017).

The usability of slider bars has been questioned in multi-device surveys (Chyung et al., 2018). Some studies highlighted that the performance of slider bars is especially poor on mobile devices (Antoun, Katz, Argueta, & Wang, 2018; Bosch et al., 2019; Funke, 2016). Buskirk, Saunders, and Michaud (2015) found that, even though smartphone respondents showed a higher appreciation of slider bar questions, they also displayed worse completion outcomes such as longer completion time and more missing data. Funke (2016) compared slider scales to radio buttons in an online survey administered to German university students and found that the slider bars produced significantly higher break-offs, especially on mobile devices (with one of the three respondents abandoning the survey). Antoun, Couper, and Conrad (2017) assessed the accuracy of reporting the age with the slider (without numeric feedback) in a Dutch probability online panel and, by checking the results against the record at disposal, found that the average error rate was higher on mobile devices. However, in terms of answer distributions, Buskirk and Andrus (2014) found no differences between PC and smartphone completion on slider bar questions in a randomized experiment in an online opt-in panel. Bosch et al. (2019) found that slider bars do not threat data quality even if smartphone respondents are included in the sample.

Differences across devices in answering on slider bars may be due to technical reasons (Antoun et al., 2018). Funke (2016) explained that the gesture needed to drag the handle is the same as the one needed for scrolling on smartphones, potentially causing respondents to involuntarily close the app and drop out of the survey. Other issues may arise from the small screen size (e.g., need to zoom), from the small font size, and/or from the disproportionate size of the finger in comparison to the size of the handle on the smartphone, which intensify the physical effort of answering.

By increasing the effort, these technical hurdles may affect also the cognitive process of answering as they make it more difficult to report the chosen score. In addition, mobile respondents allegedly pay less attention than PC ones, since they can take the survey virtually from anywhere. If task difficulty increases and attention is lower, satisficing behaviors are more likely to occur. The satisficing theory (Krosnick, 1991) suggests that respondents tend to look for shortcuts (e.g., straightlining, selecting nonsubstantive answers) to reduce the cognitive effort of answering the survey. As a result, the answering process is not performed optimally (Barge & Gehlbach, 2012; Krosnick, 1991) and measurement quality is negatively affected.

The devices on which a web survey can be taken present a large variability in terms of screen size, resolution, input method (e.g., touch screen or keyboard), and so on across and within classes of devices. For parsimony purposes, we will consider three broad types of devices: computers or PCs, tablets, and smartphones. While computers and tablets, despite large variability, are comparable in terms of screen size, smartphones tend to have smaller screens. The input method, on the other hand, varies between the computer (where keyboard and, in the case of slider bars, mouse/touchpad are used) and mobile devices (tablet and smartphone) where the finger taps directly on the screen.

Research Question and Hypotheses

This study investigates the usability of slider bars in multi-device web surveys by taking into account three slider bars–related behaviors: (a) deviating from the default position of the handle, (b) anchoring to the extremes, and (c) rounding.

The overall expectation is that the smaller the device, the lower the extent to which respondents are willing to put effort into placing the handle on the slider bar. Thanks to multiple items at disposal, we can analyze answer patterns and regularities.

An experiment varying the initial position of the handle (described in Initial Position of the Handle subsection) allowed exploring the deviation from the default position. The initial position of the handle affects the selected score under certain conditions (Liu & Conrad, 2019; Toepoel & Funke, 2018) since respondents are primed into selecting scores on a specific side of the bar, hence it is expected that:

Hypothesis 1: Respondents with the handle on the left select lower scores (closer to the left extreme) than the respondents with the handle on the right.

The items had different polarities; thus, positive and negative adjectives had been placed at each extreme, making it unlikely that a consistent selection of scores on the same side of the bar indicates a strong polarization about the topic under investigation. The effect is expected to be stronger for mobile devices as they are likely to induce satisficing behaviors due to the technical difficulties in moving the handle, discouraging further movements after the first, and leading to "settle" with a less optimal score which is likely to be found closer to the default value. Two competing hypotheses are formulated. Firstly, we attribute the effect to the small screen size expecting that:

Hypothesis 2a: The effect of the initial position of the handle on the score is stronger on smartphone respondents compared to tablet and PC respondents.

Alternatively, due to the input method, it is expected that:

Hypothesis 2b: The effect of the initial position of the handle on the score is stronger on smartphone and tablet respondents compared to PC respondents.

Technical features such as input methods and screen size may also have an impact on the tendency to select extreme scores because they make it more difficult and demanding to place the handle in a specific segment of the slider bar and therefore encourage anchoring to the extremes. Two competing hypotheses are tested in this case too. On the one hand, due to the screen size, it is expected that:

Hypothesis 3a: Smartphone respondents are more likely to select extremes compared to tablet and PC respondents.

On the other hand, however, due to the input method, it is expected that:

Hypothesis 3b: Smartphone and tablet respondents are more likely to select extremes compared to PC users.

Finally, scores ending in 0 or 5 (on a scale from 0 to 100) were considered as rounded. An experiment (described in Initial Position of the Handle subsection) consisted of displaying dynamic (i.e., appearing only when moving the handle) numeric feedback to half of the sample, whereas the other half would get no numeric feedback at all. First, it is hypothesized that:

Hypothesis 4: Rounding occurs more often when numeric labels are displayed.

However, the presence of numeric labels might have a different prominence depending on the device employed. Although the cognitive process behind rounding is considered satisficing (Gideon, Helppie-McFall, & Hsu, 2017) and should be reinforced by using smaller devices, due to our design (dynamic numeric feedback or no numeric feedback), it may be more physically demanding to select a precise, rounded, score rather than selecting any other scores around it. Hence, we expect that the smaller the device, the lower the likelihood of selecting rounded scores even when numeric labels appear on screen. There are different possible explanations, since on the one side:

Hypothesis 5a: Due to the small screen size and font of labels on smartphones, when numeric labels appear, rounding will occur more often among PC and tablet users compared to smartphone users.

On the other hand, however:

Hypothesis 5b: Due to the input method (touch screen vs. mouse), when numeric labels appear, rounding will only be more frequent among PC respondents compared to tablet and smartphone users.

Data and Methods

Data for this study have been collected at the University of Trento (Italy) in the years 2015–2016 by means of two surveys, designed within the framework of the master's course *Survey Design for Social Research* and the *Academic Careers and Professional Paths Monitoring Centre*.[1] Firstly, a web survey on students' satisfaction (SSat) about university services was administered to all the students enrolled at the University of Trento, and data were collected between February and May 2015 (see Bison, Raimondi, & Passaretta, 2015, in Italian). Of the 17,817 students contacted, 7,768 (43.6%) visited the survey web page and 6,346 answered to at least 85% of the questionnaire (35.6% of the total students contacted, 81.7% of those who opened the survey). Secondly, a web survey on the Use of Space and Time (UoST) was administered in the spring of 2016 (the report describing the survey is available upon request). Of the 15,973 students contacted, 6,024 (37.7%) opened the survey and 4,423 completed at least 80% of the questionnaire (27.7% of the whole population, 73.4% of those who opened the survey). After the listwise deletion of missing values, the samples comprised, respectively, 6,343 and 4,124 respondents.

The choice of the language of completion (Italian or English) was up to the respondent. The link to the surveys was sent to the students via their University e-mail address, together with several reminders. The choice of the device was voluntary, hence generating potential selection effects: for instance, women were consistently more likely to use a mobile device rather than a PC, while master students were generally more likely to use a PC (see Online Appendix A). For this reason, we controlled for gender, age, and type of studies in our analyses. Aside from the standard mobile optimization features of the LimeSurvey software (version 2.06) at the time of the surveys, no further measures were undertaken. In order to detect the device used, we collected User Agent Strings (UAS) at the first access to the surveys (see Anchoring subsection). No incentives were offered to respondents. Some information on the two surveys is summarized in Table 1.

Experiments

Two experiments embedded in the surveys allowed to explore sliders' features. A split ballot was applied, with students randomly assigned to one of the two experimental conditions. The SSat survey included an experiment on a slider bar question inquiring about the perception of the University website. The question consisted of one stimulus[2] and six couples of adjectives (Useful/Useless, Complicated/Simple, Dynamic/Static, Unorganized/Organized, Thorough/Lacking, and Old/New), and the scale went from 0 to 100, with 50 as the central point. In the first experimental condition, no

Table 1. Overview of Data Sets Employed.

Survey's characteristics	Students' Satisfaction (SSat)	Use of Space and Time (UoST)
Population	All students	All students
Mode	Web survey	Web survey
Period of data collection	February/April 2015	May/June 2016
Paradata (User Agent Strings)	Yes	Yes
Question on device in use	No	Yes
Software	LimeSurvey	LimeSurvey
Size of population	17,817	15,973
Opened the survey (% of population)	7,768(43.6%)	6,024(37.7%)
Completed the survey (% of population)	6,346(35.6%)	4,423(27.7%)

numeric labels appeared while the handle was moved along the horizontal line (see Figure 1 on the left). In the second condition, on the contrary, numeric labels appeared (see Figure 1 on the right) while dragging the handle.

The UoST survey contained another experiment on sliders. Although a screenshot is only available for one of the two conditions (see Online Supplemental Figure B1), the layout was the same as in the other experiment except for the initial position of the handle and the fact that no numeric labels appeared. In the first experimental condition, the handle was initially placed at the right-side extreme of the slider bar, while in the second condition, it was at the left-side extreme. The question regarded the perception of the towns of Trento/Rovereto[3] and comprised eight couples of adjectives: Happy/Sad, Close/Open, Fast/Slow, Inefficient/Efficient, Sweet/Sour, Old/New, Pleasant/Unpleasant, and Intolerant/Tolerant. In this case, the scale ranged from –10 to + 10, with 0 as the central point.

Unlike the first, in the second experiment, answers were not mandatory. Therefore, it was possible for the respondents to skip the item and leave the handle at its original position. In that case, the recorded value corresponded to the extreme (−10, or +10, depending on the experimental condition). This constitutes a limitation as it makes it impossible to distinguish nonresponses from valid, selected extreme values. In order to reduce the bias, 83 cases displaying only extreme answers on one side (hence, either −10 or + 10) and standard deviation 0 were excluded from the analyses as such consistency may indicate nonresponse. However, for the purposes of this study, extreme values can be treated as indicators of satisficing behaviors, regardless of whether they have been selected on purpose or not, as even nonresponses constitute cognitive shortcuts.

Variables

Device employed. The device employed has been recorded in two ways: In both surveys, paradata were collected to detect the device used by means of UAS. UAS are text variables "that can be captured when connecting to a website" (Callegaro, 2010, p. 3) and provide information on the device employed to connect to the Internet. Roßmann and Gummer (2014) developed a Stata program to decode the information included in the UAS.[4] In the UoST survey, additionally, students were asked directly what about the device they were using.[5] The survey was opened with a smartphone by one of three students in the SSat (28.6%) and by two of five in the UoST (42.3%; see Table 2). The use of tablets to complete surveys is far less widespread than the use of smartphones: Less than 1 of 10 students used a tablet in the SSat survey (7.4%) and even less than 1 of 20 in the UoST survey (4.6%).

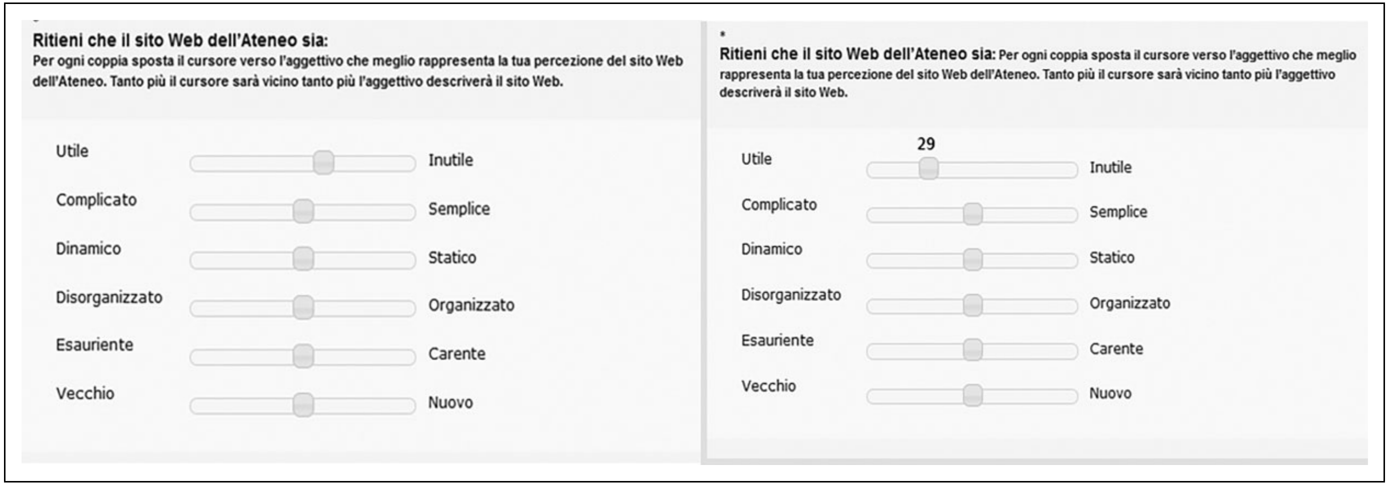

Figure 1. Screenshot of experiment on numeric labels, Conditions A (on the left) and B (on the right).

Table 2. Descriptives.

	Satisfaction Survey (2015)		Use of Space and Time (2016)	
Variables	%	*N*	%	*N*
Device used at access	*n* = 6,844		*n* = 5,555	
Desktop/laptop	63.97	4,378	53.07	2,948
Smartphone	28.62	1,959	42.29	2,349
Tablet	7.41	507	4.64	258
Device used (after listwise deletion)	*n* = 6,343		*n* = 4,124	
Desktop/laptop	65.84	4,176	58.83	2,426
Smartphone	26.77	1,698	36.42	1,502
Tablet	7.39	469	4.75	196
Control variables	*n* = 6,343		*n* = 4,124	
Sex: Female	55.71	3,534	56.01	2,310
Age: 18–20	16.16	1,025	17.00	701
Age: 21–22	28.16	1,786	34.12	1,407
Age: 23–24	24.04	1,525	24.13	995
Age: 25–26	16.35	1,037	13.36	551
Age: 27+	15.29	970	11.40	470
Type of study: Master	41.60	2,639	33.07	1,364
Selection of extremes (0–1)	*n* = 6,343		*n* = 4,124	
Extreme(useful/useless)	0.29	1,816	—	—
Extreme [complicated/simple]	0.12	747	—	—
Extreme (dynamic/static)	0.07	422	—	—
Extreme (unorganized/organized)	0.11	688	—	—
Extreme (thorough/lacking)	0.09	547	—	—
Extreme (old/new)	0.09	666	—	—
Extreme(happy/sad)	—	—	0.17	0.37
Extreme(close/open)	—	—	0.20	0.40
Extreme(fast/slow)	—	—	0.17	0.37
Extreme(inefficient/efficient)	—	—	0.28	0.45
Extreme(sweet/sour)	—	—	0.18	0.39
Extreme(old/young)	—	—	0.23	0.42
Extreme(pleasant/unpleasant)	—	—	0.24	0.43
Extreme(intolerant/tolerant)	—	—	0.20	0.40
Sliders and dependent variables	*n* = 6,343		*n* = 4,124	
	Mean	*SD*	Mean	*SD*
Useful/useless (0–100)	21.85	20.99	—	—
Complicated/simple [0–100]	57.87	26.38	—	—
Dynamic/static (0–100)	45.15	22.18	—	—
Unorganized/organized (0–100)	61.65	24.59	—	—
Thorough/lacking (0–100)	37.89	22.62	—	—
Old/new (0–100)	63.50	22.07	—	—
Happy/sad (−10–+10)	—	—	−2.21	5.40
Close/open (−10–+10)	—	—	−1.42	5.92
Fast/slow (−10–+10)	—	—	−0.98	5.42
Inefficient/efficient (−10–+10)	—	—	5.25	4.70
Sweet/sour (−10–+10)	—	—	−1.50	5.29
Old/young (−10–+10)	—	—	−2.81	5.55
Pleasant/unpleasant (−10–+10)	—	—	−4.24	4.93
Intolerant/tolerant (−10–+10)	—	—	0.97	5.89
Rounded scores (0–6)	2.05	1.67	—	—

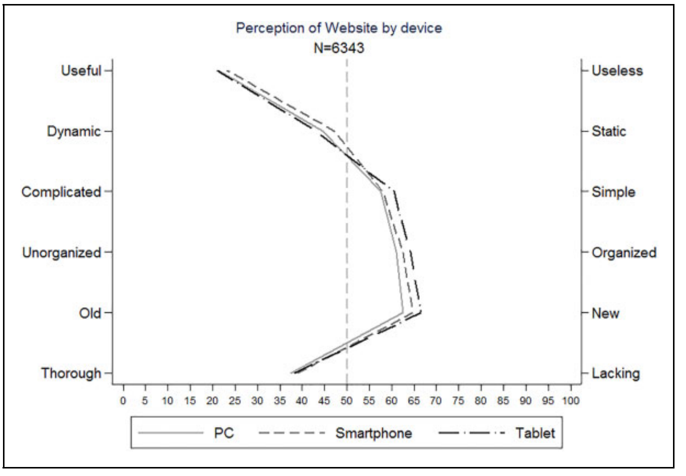

Figure 2. Average scores of semantic differentials by device (students' satisfaction). *Source*. Students' satisfaction survey (2015).

Survey completion. The surveys were considered completed (and respondents retained) if around 80% of the questionnaire had been completed. Overall, smartphone users were more likely to break off in comparison to PC and tablet users. This may be due to the lack of mobile optimization, which may have discouraged respondents. In the SSat survey, smartphone users tended to break off 3 times more than PC users; in the UoST survey, more than one of four of those who opened the survey on a smartphone broke off, which is twice the PC users. In comparison to PC users, tablet users broke off more in the SSat survey. However, in the UoST survey, the difference disappears (see Online Appendix C).[6]

Control variables. The distribution of the variables used as controls (gender, age, and type of study) is summarized in Table 2. These variables have been chosen because they seem to affect the choice of the device (see Online Appendix A).

Semantic differential. Regarding the overall scores assigned in the slider bars, the website was perceived as useful, new, and thorough. The towns of Trento/Rovereto were, on the other hand, perceived as pleasant and efficient but also rather old. The distribution of the scores varies slightly according to the device employed in the SSat survey (see Figure 2). Some differences between average scores among the three devices can also be noticed in the slider bars from the UoST survey (see Figure 3). In particular, tablet users appear to have slightly more polarized perceptions on at least three adjectives: In their opinion, the cities were younger, more tolerant, and happier. Differences between smartphone and PC users are less prominent.

The operationalization of other variables is described in Summary and Discussion section.

Findings

Initial Position of the Handle

This section explores the placement along the bar and the effect of the initial position of the handle. Figure 4 represents the average scores on the 8 items estimated on the basis of linear regression models. In line with Hypothesis 1, respondents visualizing the handle on the left show a consistent and significant negative effect on the average scores—that is, closer to the left extreme. The impact

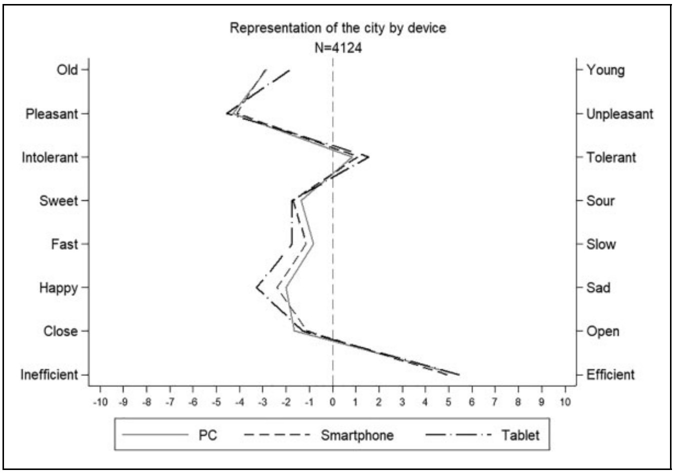

Figure 3. Average scores of semantic differentials by device (the Use of Space and Time). *Source*. Use of Space and Time (2016).

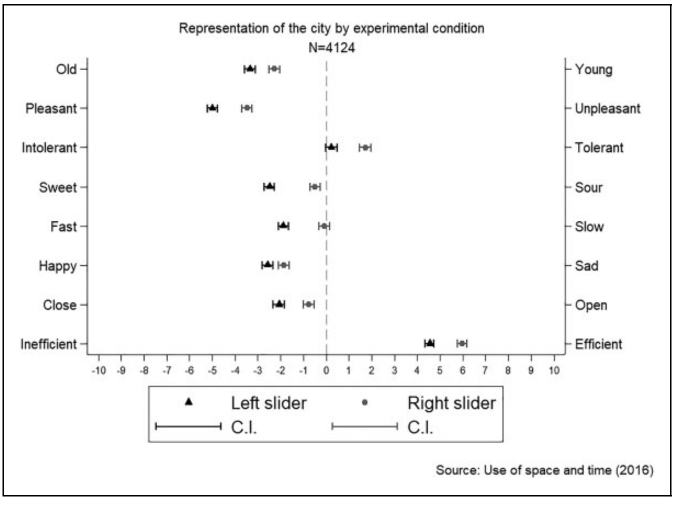

Figure 4. Predicted scores by experimental group estimated via ordinary least squares regression models (controlling for gender, age, and type of study). *Source*. Use of Space and Time (2016).

varies among items/bars in terms of size, ranging between 0.7 and 2 points of difference on average (on a scale from -10 to $+10$), but the negative effect occurs consistently.[7]

Looking at differences across devices, it appears that the negative effect of initially having the slider on the left is reinforced by the use of smartphones on all items but one. Figure 5 represents the results of a linear regression modeling the impact of experimental group, device, and their combination on the overall scores (controlling for gender, age, and type of study). Interestingly, in the first bar (happy/sad), the reinforcing effect does not occur as the use of smartphones combined with the left slider has a positive effect and, therefore, it reduces the negative impact of the initial position of the handle. It may be that the difficulties in placing the handle on the first bar trigger satisficing behaviors on the following ones, making respondents more sensitive to the priming effect due to the

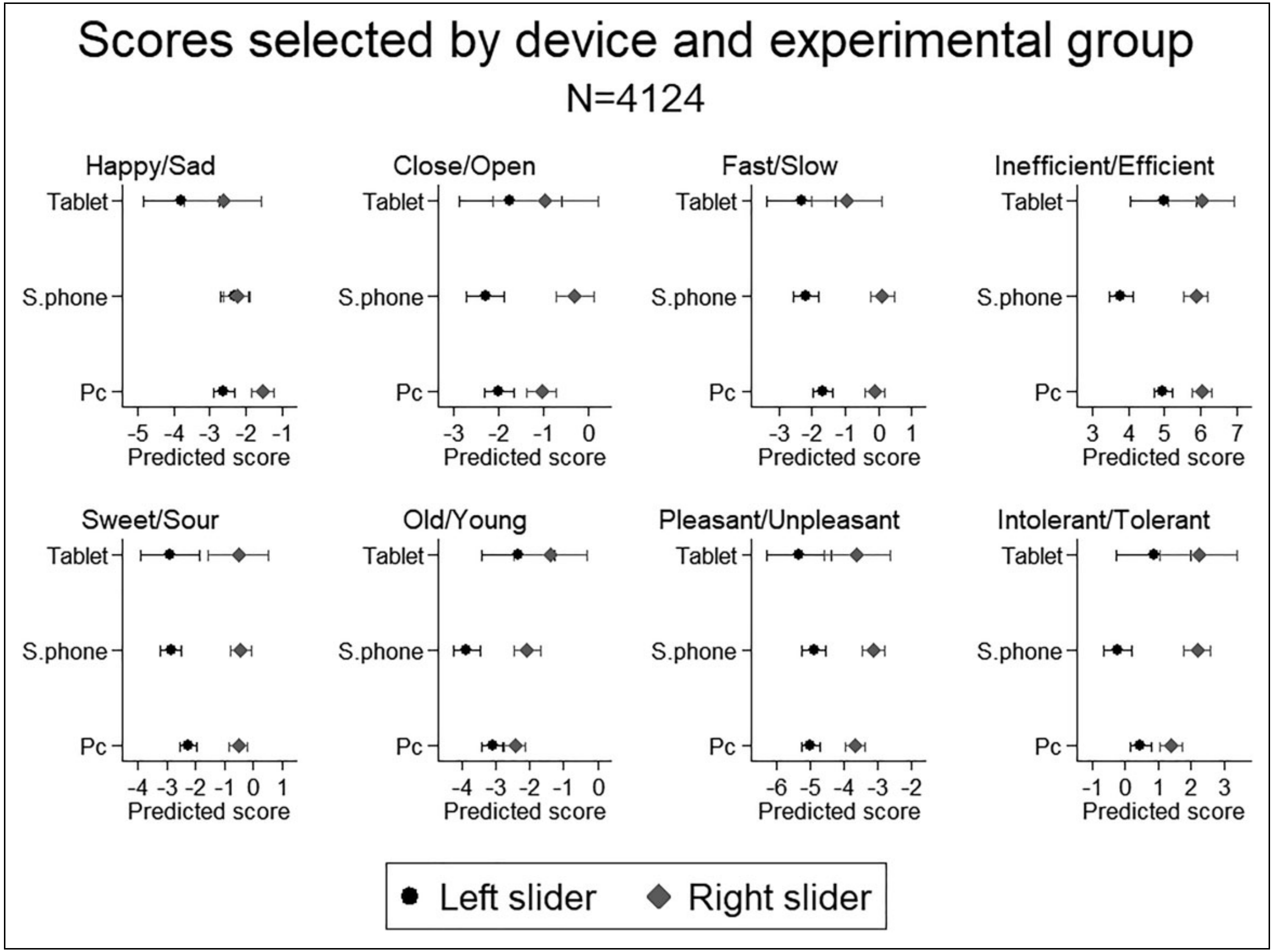

Figure 5. Predicted scores by experimental group and device estimated via ordinary least squares regression models (controlling for gender, age, and type of study). *Source*. Use of Space and Time (2016).

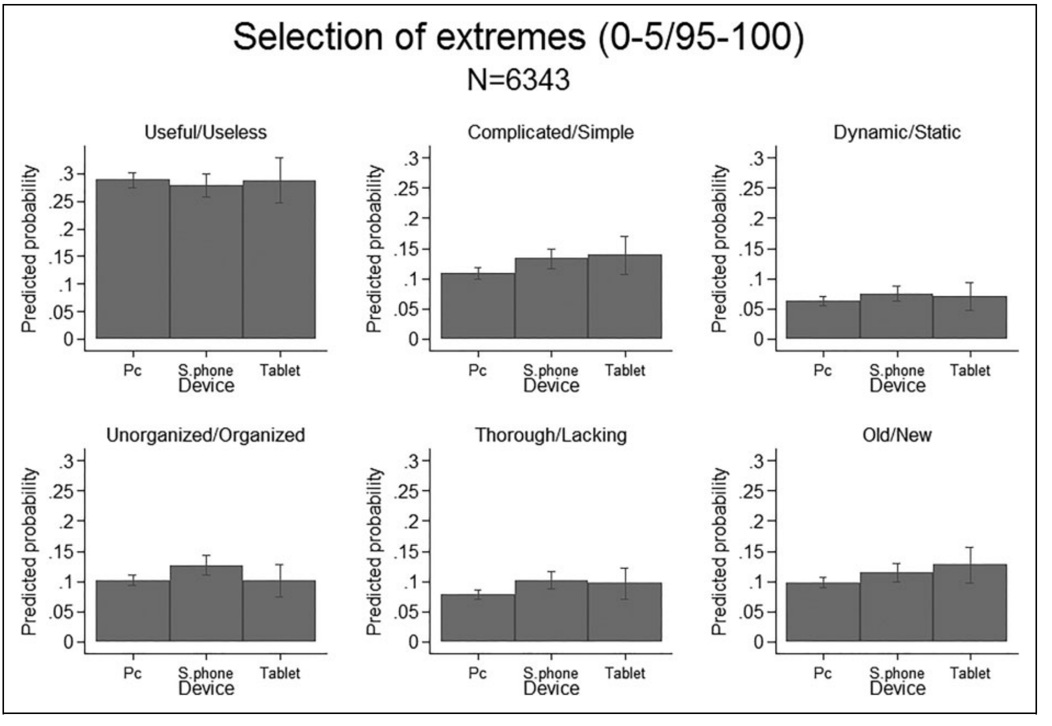

Figure 6. Predicted probabilities (estimated via binomial logistic regression) of selecting an extreme value by device (and controlled for gender, age, and type of study). *Source*. Students' satisfaction survey (2015).

initial position of the handle. In fact, not only is the main effect of the left slider negative and significant but also the interaction term of the left slider combined with the use of smartphones has a significant and negative effect on the score (the interaction term is not significant only on the 7th item, pleasant/unpleasant). As for tablets, it seems that the effect of the initial position of the handle is the same as on PCs, leading us to reject Hypothesis 2b. Results suggest that the discriminating factor lies in the size of the screen of the device employed rather than on the input method as claimed in Hypothesis 2a.

Anchoring

This section focuses on anchoring effects. In order to test the hypotheses, dummy variables were created to indicate whether respondents selected an extreme score. In the case of the UoST survey, we considered as extremes the scores -10 and $+10$. For the purpose of increasing comparability between the two scales, larger segments were considered as extremes in the SSat survey, namely scores ranging from 0 to 5 and from 95 to 100. The effect of the device on the probability of selecting extreme scores was estimated via binomial logistic regressions. Results are plotted in Figures 6 and 7.

Looking at smartphone respondents, two different situations occur. In the UoST survey, where the handle is placed at the extremes (either left or right one), smartphone users have a higher propensity—ranging from 1.8 times more in the item happy/sad to 3.1 times more on the item sweet/sour—to anchor to the extremes in comparison to PC users, and the effect is consistent on all the eight slider bars (see Figure 7). Yet, when the handle is in a central position like in the SSat survey (see Figure 6), no consistent differences in anchoring occur across devices. Except for the

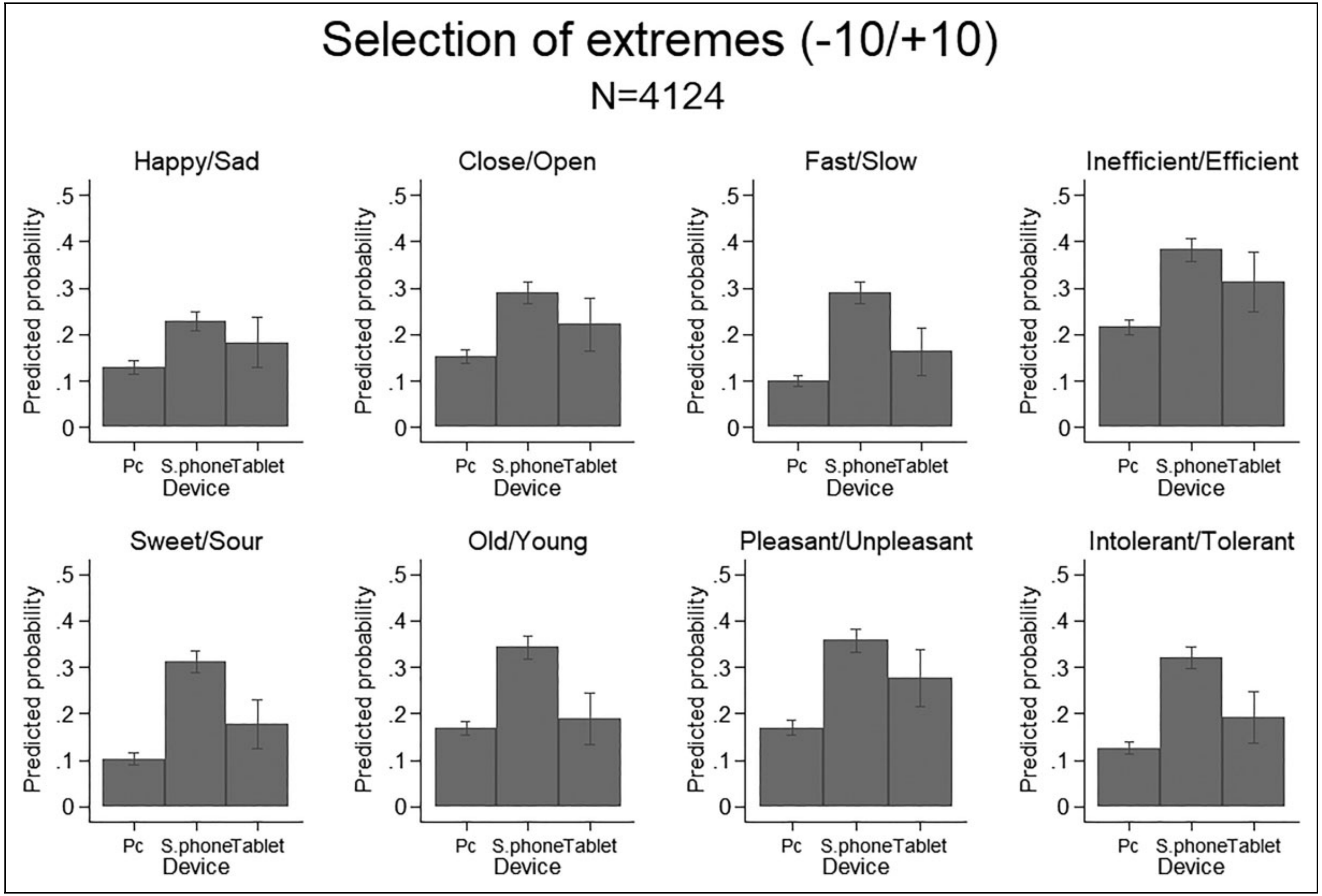

Figure 7. Predicted probabilities (estimated via binomial logistic regression) of selecting an extreme value by device (and controlled for gender, age, and type of study). *Source*. Use of Space and Time (2016).

Table 3. Linear Regression Model.

	Number of Rounded Scores Selected	
Predictors	*b*	*SE*
Constant	1.75***	.06
Experimental group: numeric labels	0.66***	.04
Female	−0.02	.04
Age (ref. 19–20)		
21–22	0.02	.04
23–24	0.04	.07
25–26	0.02	.08
27+	0.17**	.08
Master student	−0.12**	.05
R^2	0.04	

Source. Students' satisfaction survey (2015).
Note. $N = 6{,}343$.
*$p < .1$. **$p < .05$. ***$p < .01$.

items useful/useless and old/new, the effect of smartphone use on selecting extremes is positive but small on all other items. As for tablet respondents, no consistent effects are found. In fact, only in the case of the UoST survey, it looks like there may be differences from PC users and only on some of the bars.

In a nutshell, as regard anchoring effects, supportive evidence for Hypothesis 3a was only found in the UoST survey and not in the SSat one. This outcome reinforces the idea that the initial position of the handle matters for the selection of particular segments of the bar.[8] As a matter of fact, smartphone users were more likely to select extremes than PC and tablet users only when the handle was already placed at the extremes. It seems that tablet and PC respondent behave similarly, therefore discarding the idea that the input method encourages anchoring to the extremes.

Rounding

This section investigates the rounding behavior and the results of the experiment on numeric labels. The dependent variable was constructed by counting how many scores ending in 5 or 0 were selected on the slider bars: The range spans from a minimum of 0 to a maximum of 6. Linear regression models were performed to assess the impact of device and experimental group on the propensity to select rounded scores. The expectation was that rounding occurs more often when numeric labels appear. The results reported in Table 3 support Hypothesis 5. When numeric labels were visible, respondents tended to select a higher number of rounded scores: On average, of the six scores assigned, respondents visualizing the labels selected 0.7 more rounded scores than respondent who did not visualize labels.

After taking into account the device used, the effect of numeric labels on the selection of rounded scores is significantly stronger for PC users compared to smartphone users (see Figure 8). However, there is no significant difference between PC and tablet users. When no numeric labels were visible, there was not much difference across devices, whereas when labels were displayed, the behavior of respondents varied. As a matter of fact, PC users selected on average 0.5 more rounded scores than smartphone respondents when labels were visible. The screen size (and, perhaps, the size of the font of the survey objects) seem to affect the extent to which numeric labels influence the selection of the scores. The touch-screen input method (e.g., the finger dragging the handle and covering the

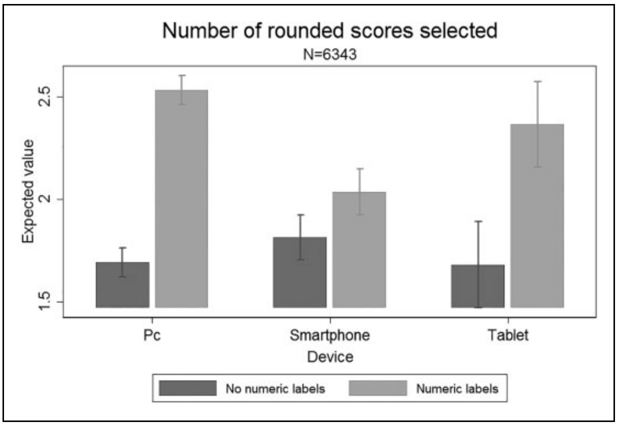

Figure 8. Predicted values (ordinary least squares regression) of number of rounded scores selected by device and experimental group (and controlled for gender, age, and type of study). *Source*. Students' satisfaction survey (2015).

numeric feedback) does not seem to be problematic in comparison to the mouse when it comes to rounded scores as tablet respondents behave like the PC ones.

Summary and Discussion

The study investigated the usability of features of sliders in multi-device web surveys. Several conclusions can be drawn from the findings. In line with previous studies, we found that the initial position of the handle matters, especially for smartphone respondents. The consistent effect, occurring on all the slider bars but the first, suggests higher satisficing among smartphone respondents as they do not move the handle as far away from its original position as other respondents. Therefore, they may be assigning a reasonable score that is the closest to the starting point rather than the most accurate score. The solution proposed by Liu and Conrad (2019) to use sliders without default values should be tested to determine whether it yields good measurement quality on smartphones too.

Results also showed that when numeric labels were visible, rounding occurred more often. However, numeric labels had a smaller impact on scores on smartphones' screens compared to bigger devices' screens. This result clearly questions the usability of sliders for accurate measures (such as, for instance, age or financial quantities) in multi-device surveys in line with previous studies (Antoun et al., 2017). Further research is needed to establish the impact of different positions, fonts, and sizes of the numeric feedback which may be more suitable for small screens. Moreover, rounding on slider bars is a cognitive shortcut (satisficing) but also an additional physical effort: Research comparing slider bars to other question types (such as text boxes or numbered VAS) may shed light on this paradox.

Results on anchoring are, finally, inconclusive. In the analyses on the UoST survey, smartphone respondents resulted more likely to select extremes, hence suggesting higher satisficing; however, in the SSat survey, differences were less clear-cut. There are two possible explanations for the differential effect. From a design point of view, placement along the bar is affected by the initial position of the handle as seen previously. From a substantive point of view, in addition, it may be that the question associated with the slider bars in the UoST was more sensitive than the question used in the other survey and, therefore, it may have stimulated more polarized answers.

On the whole, we found evidence that design choices affect answers on the slider bars: The number of rounded scores varies with or without numeric feedback, and the initial position of the handle affects scores. This implies that slider bars have to be employed carefully, and their features should always be made explicit when presenting results as they may bias scores. Moreover, we may conclude that results are only slightly leaning toward higher satisficing among smartphone respondents, while tablet and PC respondents behave similarly. Nevertheless, it was shown that features of slider bars had differential impact depending on the screen size. Hence, the usability of sliders has to be always questioned when designing a survey for multi-device completion, and extensive pretesting on different devices could be useful to detect potential problems.

Some remarks are needed to frame these conclusions. First of all, the analyses are based on a population of university students. These respondents are likely to have better technological skills compared to the general population, making generalization risky. However, this characteristic of the respondents also allowed controlling for sources of variability which may otherwise inflate the results. It is striking that differences in completion across devices emerge even in such a specialized population, which is socialized to use Internet-enabled devices. This may indicate that differences lie in a different answering process rather than in technical problems. Yet, it should also be pointed out that we were not able to fully account for sociodemographic differences across device groups, which may play a role too, especially when looking at tablet users. Further research should focus on this aspect.

Secondly, there is a strong normative assumption behind the concept of satisficing. It is often assumed that "less effort" in answering indicates lower quality answers. However, there are some cases—such as semantic differentials—where a prompt and instinctive reaction to the stimulus may be desirable.

Another limitation of this study is that the web surveys themselves were not specifically optimized for mobile devices, whereas nowadays many studies adopt mobile-first designs that allegedly mitigate differences across devices. Nevertheless, since respondents appeared to react differently to sliders' features depending on the device used, it is important to test ways of optimizing slider bars for mobile devices which will not bias results.

Finally, completion outcomes are problematic. As long as smartphone respondents drop out more—as highlighted by other studies—compared to respondents using other devices, satisficing effects are probably underestimated because technical difficulties are compensated for by the stronger motivation of those who keep answering.

Data Availability

The data and syntax used for the study are available free of charge and in compliance with the GDPR upon request to Prof. Ivano Bison (ivano.bison@unitn.it) and after obtaining authorization from the University of Trento.

Authors' note

We are grateful to the anonymous reviewers for the constructive comments on the manuscript. The surveys were collected with the technical support of E. Loner and C. Santinello, and the practical support of E. Raimondi, G. Passaretta, M. Lugo, and the students of the course in Survey Design for Social Research and of the Academic Careers and Professional Paths Monitoring Centre at the University of Trento, A.Y. 2014/2015 and 2015/2016. We are also thankful to G. Brandolini for contributing to the readability of the paper.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Software Information

Stata Version 14 has been used to perform the analyses. The following user-written packages were used:

- parseuas (by Roßmann & Gummer, 2014)
- grc1leg (by Wiggins, 2003)
- coefplot (by Jann, 2013)
- parmest (by Roger Newson)
- outreg (by John Luke Gallup)

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Both surveys have been ideated and coordinated by Prof. Ivano Bison.
2. The question is “Do you think that the University website is: (line break) For each pair of adjectives, move the cursor towards the one that best represents your opinion of the University website. The closer the cursor is to the adjective, the more it describes the website” (see also Figure 1).
3. The question is “Describe your perception of the city of [Trento/Rovereto] through the following list of adjectives in pairs. For each couple, move cursor towards the adjective which describes better the city. The closer the cursor is to one of the two adjectives, the better that adjective describes the city.”
4. In both surveys, some respondents disabled cookies (12% and 8%, respectively): For them, it was not possible to collect the User Agent Strings (UAS). Since the amount of students with disabled cookies that reached the end of the surveys is lower than 1%, it does not seem problematic to exclude them from the analysis.
5. Interestingly enough, in the survey in which both the information were available (UAS and self-reported device), a small amount of respondents (4.75% of those who completed the survey) showed discrepancies between the two pieces of information. Further tests suggested that the discrepancy might be due to a switch in the device: The UAS is recorded when the survey is opened, while the question about the device is asked at the end. Since for the purposes of the analysis, it is important to know which device has been used on each specific question, and since there was no way to detect when the switch occurred, respondents with discrepancies have been excluded from the analyses.
6. The average interview time is slightly higher for smartphone users in comparison to other devices’ users, while there is little difference between PC and tablet.
7. Even when computing the mean score assigned to the eight sliders for each individual, those with the right slider assigned on average 1.41 points more than those with the left slider. The difference becomes significantly larger on smartphones (1.81) compared to PC (1.08) and tablet (1.34).
8. We also explored the anchoring behavior to center values (45–55 in the students’ satisfaction [SSat] survey, −1 to +1 in the Use of Space and Time [UoST] survey), expecting that in the SSat survey, where the handle is initially at the center, smartphone users are more likely to select central scores. Results are reported in Online Appendix D. Results presented in Online Supplemental Figure D1 show some significant but small differences across devices in anchoring to central scores. However, the direction is the opposite of the one expected. Smartphone and tablet respondents are less likely to select central scores, even though the effect is—as said—small. Similarly, in the UoST survey, completing the survey with the smartphone has a negative and mostly significant effect on selecting central scores (−1 to +1). In some cases, such as the items fast/slow or sweet/sour, the difference between PC and smartphone respondents is clearly marked (see Online Supplemental Figure D2).

References

Antoun, C., Couper, M. P., & Conrad, F. G. (2017). Effects of mobile versus PC web on survey response quality. *Public Opinion Quarterly*, *81*, 280–306. doi:10.1093/poq/nfw088

Antoun, C., Katz, J., Argueta, J., & Wang, L. (2018). Design heuristics for effective smartphone questionnaires. *Social Science Computer Review*, *36*, 557–574. doi:10.1177/0894439317727072

Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, *53*, 182–200.

Bison, I., Raimondi, E., & Passaretta, G. (2015). *Indagine sull'uso e la soddisfazione dei servizi per gli studenti dell'Universitá di Trento* [Survey on the use and satisfaction of the services for the students of the University of Trento]. Trento, Italy. Retrieved from https://webmagazine.unitn.it/news/ateneo/9633/indagine-sull-uso-dello-spazio-e-del-tempo-degli-studenti-di-unitrento

Bosch, O. J., Revilla, M., Decastellarnau, A., & Weber, W. (2019). Measurement reliability, validity, and quality of slider versus radio button scales in an online probability-based panel in Norway. *Social Science Computer Review*, *37*, 119–132. doi:10.1177/0894439317750089

Buskirk, T. D., & Andrus, C. (2014). Making mobile browser surveys smarter: Results from a randomized experiment comparing online surveys completed via computer or smartphone. *Field Methods*, *26*, 322–342. doi:10.1177/1525822x14526146

Buskirk, T. D., Saunders, T., & Michaud, J. (2015). Are sliders too slick for surveys? An experiment comparing slider and radio button scales for smartphone, tablet and computer based surveys. *Methods, Data, Analyses*, *9*, 229–260. doi:10.12758/mda.2015.013

Callegaro, M. (2010). Do you know which device your respondent has used to take your online survey? *Survey Practice*, *3*, 6.

Chyung, S. Y. Y., Swanson, I., Roberts, K., & Hankinson, A. (2018). Evidence-based survey design: The use of continuous rating scales in surveys. *Performance Improvement*, *57*, 38–48. doi:10.1002/pfi.21763

Cook, C., Heath, F., Thompson, R. L., & Thompson, B. (2001). Score reliability in web- or internet-based surveys: Unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement*, *61*, 697–706. doi:10.1177/00131640121971356

Couper, M. P., Antoun, C., & Mavletova, A. (2017, January 27). Mobile web surveys. In P. P. Biemer, E. D. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, ... B. T. West (Eds.), *Total survey error in practice* (pp. 133–154). New York, NY: Wiley. doi:10.1002/9781119041702.ch7

Couper, M. P., Tourangeau, R., Conrad, F. G., & Singer, E. (2006). Evaluating the effectiveness of visual analog scales: A web experiment. *Social Science Computer Review*, *24*, 227–245.

De Bruijne, M., & Wijnant, A. (2014). Mobile response in web panels. *Social Science Computer Review*, *32*, 728–742.

DeCastellarnau, A. (2018). A classification of response scale characteristics that affect data quality: A literature review. *Quality and Quantity*, *52*, 1523–1559. doi:10.1007/s11135-017-0533-4

Funke, F. (2016). A web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales. *Social Science Computer Review*, *34*, 244–254. doi:10.1177/0894439315575477

Funke, F., & Reips, U.-D. (2012). Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales. *Field Methods*, *24*, 310–327. doi:10.1177/1525822X12444061

Funke, F., Reips, U.-D., & Thomas, R. K. (2011). Sliders for the smart: Type of rating scale on the web interacts with educational level. *Social Science Computer Review*, *29*, 221–231. doi:10.1177/0894439310376896

Gideon, M., Helppie-McFall, B., & Hsu, J. W. (2017). Heaping at round numbers on financial questions: The role of satisficing. *Survey Research Methods*, *11*, 395–404. doi:10.1021/acssynbio.5b00266

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*, 213–236.

Liu, M., & Conrad, F. G. (2016). An experiment testing six formats of 101-point rating scales. *Computers in Human Behavior*, *55*, 364–371. doi:10.1016/j.chb.2015.09.036

Liu, M., & Conrad, F. G. (2019). Where should I start? On default values for slider questions in web surveys. *Social Science Computer Review*, *37*, 248–269. doi:10.1177/0894439318755336

Maloshonok, N., & Terentev, E. (2016). The impact of visual design and response formats on data quality in a web survey of MOOC students. *Computers in Human Behavior*, *62*, 506–515. doi:10.1016/j.chb.2016.04.025

Roßmann, J., & Gummer, T. (2014). PARSEUAS: Stata module to extract detailed information from user agent strings. *Statistical Software Components*. Retrieved from https://ideas.repec.org/c/boc/bocode/s457937.html

Roster, C. A., Lucianetti, L., & Albaum, G. (2015). Exploring slider vs. categorical response formats in web-based surveys. *Journal of Research Practice*, *11*, 1–19.

Toepoel, V., & Funke, F. (2018). Sliders, visual analogue scales, or buttons: Influence of formats and scales in mobile and desktop surveys. *Mathematical Population Studies*, *25*, 112–122. doi:10.1080/08898480.2018.1439245

Author Biographies

Angelica M. Maineri is a junior researcher/PhD candidate at the Department of Sociology, Tilburg University, The Netherlands. She is also an assistant of the Methodology Group and member of the Operational and Planning Group of the European Values Study. E-mail: a.m.maineri@uvt.nl

Ivano Bison is an associate professor at the Department of Sociology and Social Research, University of Trento, Italy. His research interests focus on, but are not limited to, survey design and digital social data. E-mail: ivano.bison@unitn.it

Ruud Luijkx is an associate professor at the Department of Sociology, Tilburg University, The Netherlands, and a visiting professor at the University of Trento, Italy. He is currently the Chair of the Methodology Group of the European Values Study. E-mail: r.luijkx@uvt.nl